Urban Traffic Description: Data and Statistical Models

Il traffico urbano: informazioni e modelli statistici

Alessandro Viviani, Alessandra Petrucci
Università di Firenze, Dipartimento di Statistica “G. Parenti”
Via Morgagni, 59 Firenze
viviani@ds.unifi.it , alex@ds.unifi.it

Riassunto: L’obiettivo che ci siamo proposti in questo lavoro è stato quello di riflettere sul ruolo della statistica nell’analisi di un fenomeno complesso quale la mobilità a scala territoriale, sia per quanto riguarda natura e caratteristiche dell’ informazione sia per l’impiego di modelli.
La disponibilità di informazioni pertinenti è modesta; infatti, negli approcci allo studio della mobilità si sono verificati solo dei tentativi di ricomposizione di fonti sistematiche (statistiche e non) e di raccolte periodiche di dati.
In questa direzione, ci siamo proposti di far riferimento al dato censuario ed alla sua utilizzazione nelle cosiddette matrici Origine/Destinazione con l’applicazione di un modello d’interazione spaziale per l’analisi dei flussi a livello territoriale; questo rappresenta un caso di trattamento di informazioni statistiche (definite, raccolte e trattate, cioè, con finalità statistiche) mediante l’impiego di un modello specificato e stimato allo scopo di individuare comportamenti collettivi. Questo allo scopo di fornire una “misura” del fenomeno, almeno a livello urbano.
In secondo luogo, ci siamo posti l’obiettivo di esaminare un rilevante attributo della mobilità, la sicurezza stradale; questa, che viene spesso identificata nelle caratteristiche geometriche e strutturali delle strade viene riconsiderata attraverso l’esplorazione e l’impiego di una fonte tipicamente amministrativa (i dati sull’incidentalità rilevati dall’autorità di polizia) per ricavare sia i contenuti definitori della sicurezza stradale sia una misura della stessa.

Keywords: Spatial interaction models, accidents, geocoding, mobility

1. Introduction

Who has been involved in studying economic and social phenomena on a territorial scale knows that only the combination of data from census, direct surveys (either sample surveys or other) and administrative archives allows to build an adequate information system to analyze the structure and dynamics of such a complex phenomenon as the territorial mobility.
The combination of different sources with a view to analyzing economic and national accountancy issues has already reached a good level in Italy, especially thanks to the considerable impetus given by our late lamented friend and colleague Marco Martini.
In this work, the analysis that we perform has a limited horizon due to the scarcity of pertinent information, in terms of informative content meeting cognitive needs, on the mobility phenomenon, whereas there is a variety of data collections originated from both administrative requirements and specific project questions. Once we have excluded to present a critical review of sources on the mobility theme and their use in behavior

models derived from the literature (largely pervaded by an engineering approach), we have focused on the objective to think about the role of Statistics in the analysis of the mobility phenomenon, with respect to both the nature and characteristics of information and the use of models.
To finalize this work one of the authors refers to a personal experience at the beginning of their work at the University as a grant-holder: IRPET (the Tuscany Regional Institute for Economic Programmation) entrusted them with developing a data analysis project on the commuting (Origin/Destination) movements for work and study purposes, with a view to find out significant regularities in these movements at a territorial level. Naturally the census data allow (as it allowed at that time) to make "structural" observations of the phenomenon and long-term analyses in relation to the "long" times of the dynamics between two census periods, whereas fast changes require "short" times that overlap and intersect in census intervals. If we add to this the (unfortunately structural too) delays in processing census data, we better understand the need for information characterized by accuracy, confrontability, coherence, completeness, relevance and, especially, timeliness.
For these reasons attention was drawn towards the use of data from administrative sources; however the interest in these data did not lead to the systematization that happened in the field of business economic statistics. Among the various approaches to the mobility phenomenon, there have been only attempts to recompose, at different levels, systematic sources (statistical and not statistical) and periodical data collections relating to specific projects or coming from the innovation of data capturing technological supports, supports that have often been presented as sources of information rather than supports for data surveys. Thus there is an overlap of data obtained from urban planning studies, data on transport demand and supply (usually split by means of transportation) and the so-called indirect sources (data on pollution, fuel consumption, car accidents...), but a coherent system, shared by the different agents interacting in this field, has not been developed. The above-mentioned observations explain why the hypothesis of a critical review of sources and their necessary combination, possible at a very fine territorial level, has been given up, and new attention has been turned on two aspects, specific but of general interest, that are also products of the collaboration among the Florence Municipality and the Department of Statistics "G. Parenti" at the University of Florence.
First we intend to examine the fundamental problem – at least for a statistician – of the mobility phenomenon, i.e. its measurement. To this end we can only refer to the census data and its use in the so-called Origin/Destination matrices when applying a spatial interaction model for the analysis of flows at a territorial level. This is a case of treatment of statistical information (i.e. the information is defined, collected and processed for statistical purposes) through the use of a model specified and assessed with a view to characterize a collective behavior.
Second, we intend to take the measure of road security, a relevant attribute of the mobility phenomenon. Road security, which is often identified with the geometrical and structural characteristics of the streets (width, gradient, curve radius), is here re-considered through the exploration and use of a typically administrative source (the data on accidents gathered by the police) to draw out both the defining contents and a measure of road security.

The work is structured as follows: the second paragraph deals with spatial interaction models and presents their relative application to the case study. The third paragraph introduces the accident issue in urban area, providing the results obtained from the analysis of data relating to accidents and their position on the road graph, collected in the Florence urban area. Finally, conclusions are provided in the last paragraph.

2. Spatial interaction models

The data considered in this paragraph are the journeys-to-work[1] inside the city of Florence at the census tracts level collected during the last Italian Population Census in 2001. This "flowing" data between "origins" and "destinations" are typical spatial interaction data. The primary objective in the analysis of interaction data is to understand or model the pattern of flows. If we can model such data we might then be in a position to use such a model for planning purposes. We need to emphasise the is not the primary concern to understand flows at the individual level. Rather, we want to model aggregate interaction pattern. Our interest lies in examining the spatial behaviour of collections of individuals, rather than individuals themselves. Typically, we can recognise three important sets of factor to explain such aggregate spatial interaction. The first one deals with the spatial separation of origin from destinations. The second consider the characteristics which determine the volume of flow from each origin and the last one regard the characteristics of the destination which relate to their "attractiveness".

The basic structure of the spatial interaction problem, then, is to express the volume of flow in term of origin, destination and spatial separation factors. In general, we will interested in models of the form:

$$Y_{ij} = \mu_{ij} + \varepsilon_{ij}$$

where, $E(Y_{ij}) = \mu_{ij}$ and ε_{ij} is an error about the mean. The methodology for spatial interaction data develop suitable models for μ_{ij} involving parameters which reflects characteristics of the origin *i* (relating to the propensity to flow from each origin); characteristics of the destination *j* (relating to their attractiveness); and the deterrent effect of the distance d_{ij} between origin *i* and destination *j*.

One of the most common classes of models for μ_{ij}, which simultaneously incorporate the effect of origin and destination characteristics as well as distance, are known as gravity models. The term arise because of their analogy with Newton's law of gravity, where force of attraction is proportional to square of the distance between them.

Following the entropy maximization approach, μ_{ij} can be modelled in the general form:

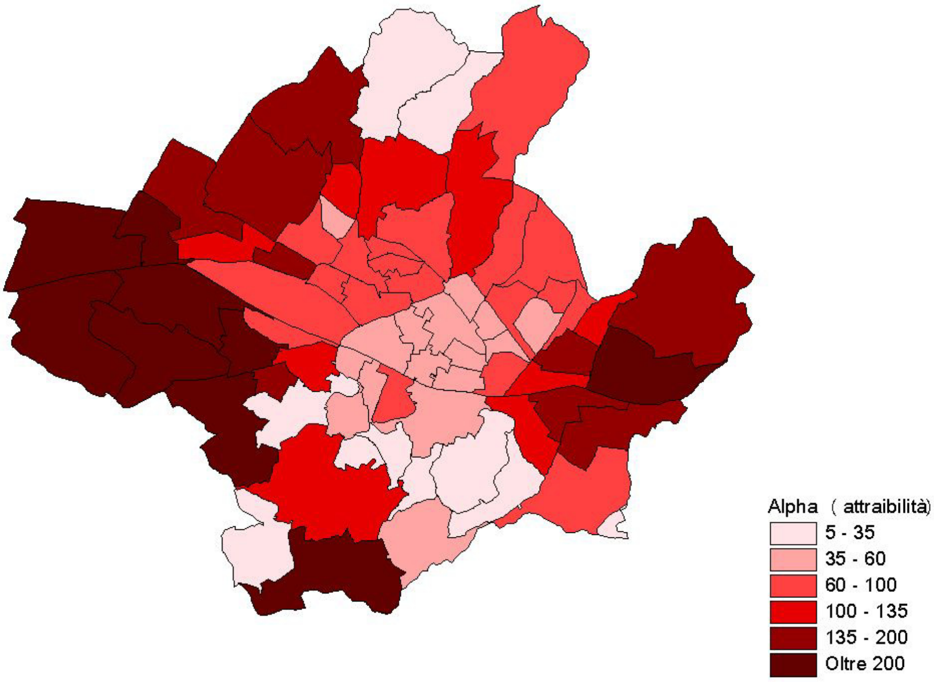
$$\mu_{ij} = \alpha_i \beta_j e^{\gamma d_{ij}}$$

where $\alpha_i = e^{-\lambda_i^{(o)}}$, $\beta_j = e^{-\lambda_j^{(d)}}$, and $\gamma = -\lambda$ simply re-express the Lagrange multipliers in a more convenient form. In this model, α_i are interpreted as a set of parameters which characterise the propensity of each origin to generate flows, β_j a set of parameters

[1] Referred to workers and students

which characterise the attractiveness of each destination and γ a distance deterrent effect.

Figure 1 *α_i parameters of doubly constrained model for the municipality of Florence (Population Census 2001, Istat - provisional data)*



This general form of the gravity model is often referred to as a doubly constrained model. We have allowed α_i and β_j to be separate parameters for each origin and each destination and have not attempt to explain their value, simply to estimate them in such a way that predicted flows reproduce exactly the total observed flow from each origin and to each destination. In a sense this model is purely descriptive rather then offering an explanation for the observed flows but modifications could be done to allow to incorporate more explanation into α_i and β_j. However, such doubly constrained model do offer one element of explanation relative to observed flows. The distance parameter γ quantifies the relative ability of distance to deter spatial interaction, and this is a useful information. The model also allows relative comparison of the generating potential of origin, or the pull of destinations, after due allowance has been made for distance effects. We can also use this model to predict what flows will result if certain changes are made in the system.

The data considered refer to the movements that people resident in the Florence Municipality make every day to reach their places of study or work, surveyed in the 14th General Census of Population and houses in 2001.

It must be noticed that these data are provisional therefore this work proposes just a first analysis of the internal commuting phenomenon. A deeper and more complete study will be performed when the definitive data are available.

Generally speaking, the word “commuting” refer to the phenomenon of the daily movement from a place of departure (home, housing of habitual dwelling) to a place of arrival (place of work or study). The data consider exclusively the movements performed by people dwelling in Florence to reach their places of study or work in Florence and going back home every day. It must be recalled that, of the commuting movements, only the movement from home to place of work (or study) has been surveyed, and not the coming-back movement. The data also involve characteristic variables of commuting such as means and times for moving (means of transportation, time spent, hour of departure from home) relying on the answers given by those who commuted to their usual places of study or work on the Wednesday preceding the date of reference for the survey.

However socio-demographic data such as age, gender, qualification, economic sector of activity for employed people, type of school for students, are not available yet, and these data will make it possible to outline a profile of the internal commuting of Florence city dwellers for reasons of work or study.

For the calculation and analysis of distances, we have considered a reference territorial basis of 72 areas defined on the basis of the 2001 census sections.

Figures 1 and 2 provide the maps according to parameters and for the 72 areas. The value of the parameter is not high (-0.567) but there is a friction effect produced by the distance on the movements. Such a result was foreseeable since this is an application at urban territorial level and distance is not a strong deterrent to move within the city limits.

Figure 2 *β_j parameters of doubly constrained model for the municipality of Florence (Population Census 2001, Istat - provisional data)*

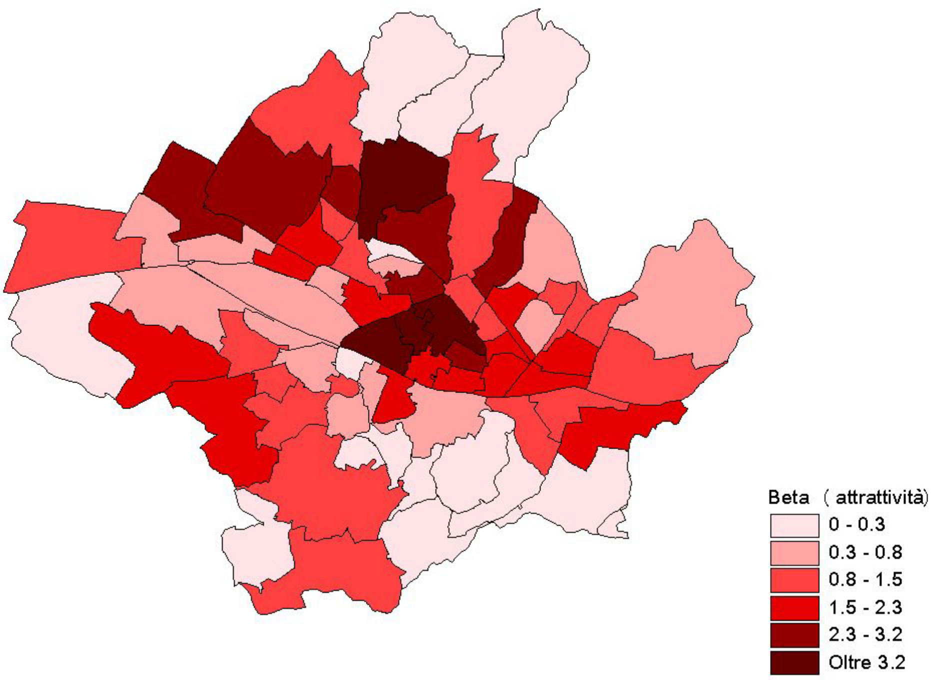


Figure 1, which represents values, shows the propensity of each zone to generate flows. We note that the areas with commuting movements are mainly those situated in the North and North-West of Florence, characterized by a great number of housing settlements, even recently built. Another area showing a propensity to generate flows is the South Eastern zone of Florence (Coverciano, Gavinana and Galluzzo) which presents characteristics similar to those of the northern and western zones of Florence.

Figure 3 *α_i parameters of doubly constrained model for the hinterland of Florence (Population Census 2001, Istat - provisional data)*

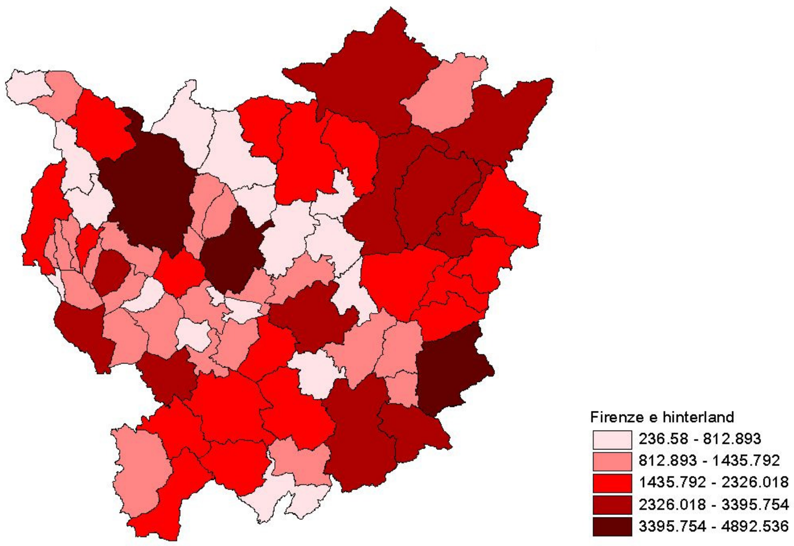
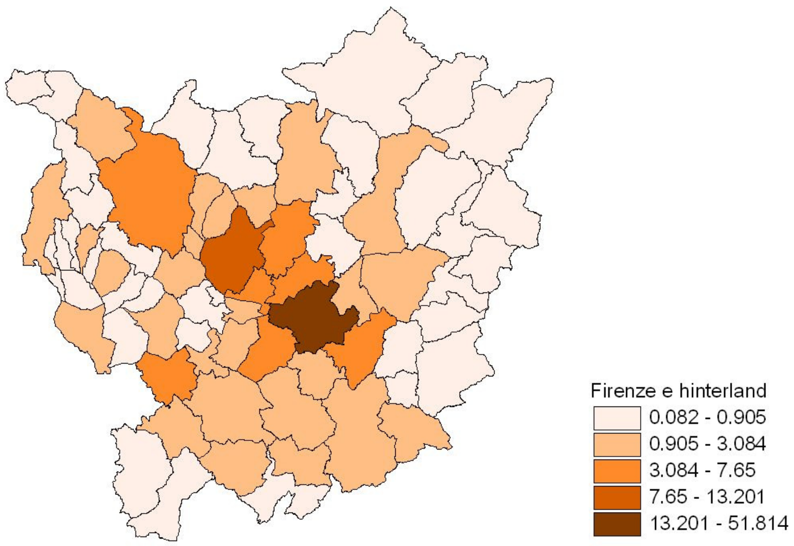


Figure 2 represents the value of parameters that express the attractiveness of each destination. In this case we note that the central areas are the ones which attract the greater quantity of flows, since offices, shops and more generally commercial activities and numerous schools, along with many university structures, are situated in these areas. Obviously the lowest values correspond to the outlying areas which generate the flows.

Figure 4 *β_j parameters of doubly constrained model for the hinterland of Florence (Population Census 2001, Istat - provisional data)*



In brief, through the parameters it is possible to know respectively the zones which have a propensity to produce flows and the ones which instead attract flows, whereas the parameter represents the relative friction effect induced by the distance.
The model obtained seems to deals satisfactorily with the process in evolution and is adaptable to the local reality to which it is applied. It can be noted that the central zones corresponding to the northern ring roads of the city attract the commuting flows, whereas the outlying zones are at their origin.
Coherent results are also obtained by modifying the territorial reading scale of the phenomenon. In fact, considering a territorial basis of 116 zones, we find the same macro areas that generate flows; however the further geographical level allows to highlight a greater specification of the zones of origin.
Moreover, to complete the descriptive analysis of urban mobility through interaction models, the parameters of the spatial interaction models have been estimated for the various means of transportation (car, motorbike/motorcycle, public transport, ecological transport) and for each reason for moving (work or study). Eventually, in order to start a study that would take into account the greater complexity of the commuting phenomenon and consider a more realistic geographical dimension, a doubly-linked spatial interaction model was implemented including also the census data relating to the commuting movements within the metropolitan area, considering as reference territorial units the municipalities wich make part of it. The parameters of this last model are reported in Figures 3 and 4.

The value of γ (-0.22), aligned with other case-studies of similar dimensions known in the literature, suggests that distance does not seem to be, at that scale, an excessive hindrance to moving.
The parameter distribution indicates that the Mugello, the South and South East of Florence and the cities of Prato and Pistoia, are the areas with the greater propensity to generate flows of movements for work and study, whereas the distribution of the corresponding parameters shows how Florence and the axis made by the three chief towns of the metropolitan area (Florence-Prato-Pistoia) constitute the pole of attraction of these flows.

3. Accidents

Over the last years, the road accident phenomenon has become more and more critical, calling the attention and involvement of the whole community, especially the ones of the authorities in charge of the intervention policies.
As a matter of fact, the constantly increasing mobility has produced heavy consequences in terms of transport security, road accidents being their most negative manifestation, becoming one of the most considerable public health issue in our country.
In conformity with the European Union Directives, Italy is bound to reduce by 40% the number of dead and injured in road accidents by the year 2010. An objective that is already pursued by implementing measures towards a major respect of road security regulations, as testify the laws on compulsory safety devices and the recent introduction of a decreasing score system applicable to the driving licence (“patente a punti”), which is a clearly a tool capable of preventing the most dangerous driving behaviour.

As in the rest of Italy, in Tuscany too the considerable mobility growth registered over the last decade has caused an increase in the number of road accidents and, although their gravity in terms of number of dead has in fact decreased, the number of injured people has been constantly rising.
The confrontation between the Tuscan chief town and other Italian cities highlights a particularly critical situation, since in Florence the road risk level is so high that it also affects the pedestrian safety. As already recalled and as it generally happens in urban areas, the accident frequency is particularly high with respect to other urban configurations although accidents have lighter consequences. In fact, in 2001 about 24% of the total number of accidents in Tuscany took place in the city of Florence whereas about 6.5% of the total of lethal accidents in the region occurred in the chief town.
Currently, as happens in other European countries, the statistical information on accidents is collected by the National Statistical Institute ISTAT, in collaboration with the Italian Automobile Club ACI, through a monthly survey of the total number of accidents involving personal injuries (either death or physical damages) on the national territory. The survey is performed through a special ISTAT form called “road accidents” to be filled in by the policemen (Road Police, Carabinieri, Municipality Police) for each accident involving a vehicle circulating on the road network and causing personal injuries.
The current information base, designed to provide its - mainly public - users with an overall set of data on road accidents, indicates that the majority of accidents and injuries occur in urban contexts, which increases the interest in studying the specific local realities. This explains the interest and necessity for the municipal authorities, in particular for the Florence Municipality, to enquire into the characteristics of this phenomenon within the limits of their jurisdiction, in order to get indications and suggestions useful to define the measures of utmost importance.
The data that we present here concern the road accident phenomenon at municipal level in the Tuscan chief town between 1995 and 2002. They are the starting point for further and deeper analyses and were collected in the framework of the project “Urban Mobility and Road Accidents” which makes part of the research programme agreement between the Department of Statistics “G. Parenti” at the University of Florence and the Florence Municipality.
Through filing and geocoding the road accidents in a GIS (Geographical Information System), we obtained a complete map of the potentially localizable accidents. This has made possible to locate those areas of circulation which prove particularly at risk for road accidents, providing the most accurate information at the territorial level: from the location of the road segments registering the higher number of accidents with personal injuries (death and physical damages) in the reference period, to the precise localization of the accidents, drawing within the limits of the available information a map of the major blacklisted roads in the municipal territory network.
In the reference period, from 1 January 1995 to 31 December 2002, there were in the Florence municipal territory 44,032 accidents, from 5,001 in 1995 to 5,577 in 2002, with an average of 5,504 accidents by year. Filing these accidents according to the consequences on persons, we observe that only 10,141 (making 23% of all accidents) involved only goods damages whereas, among those with personal injuries, 233 (0.5% of total) were lethal, and 33,658 (76.5%) resulted only in injuries. We distinguished in the analysis between accidents with deaths and injuries and accidents with only injuries, considering the first as lethal accidents: there were in total 63, representing 27% of all

lethal accidents. The number of injured persons in this category was calculated separately, as shown in Table 1.

Each year, on average, 30 persons lose their lives and 5,155 get injured. Generally speaking, whereas the total number of accidents and the number of accidents with injuries were each year higher than in 1995, the number of lethal accidents followed a different course, especially in 1998: taking as a reference the first year in the series (1995=100), we built the index numbers relating to the total number of accidents and to the total number of accidents with deaths and injuries (Table 2).

It must be underlined that all the results of our analysis refer to accident surveys made by the Municipal Police, therefore those accidents for which there was no police intervention are excluded. Given that these are generally accidents of lesser importance, they do not involve physical damages and are inevitably underestimated in this survey.

Table 1 *Accidents, deaths and injures per year*.

Year	Accidents	Accidents with injures		Fatal accidents		
		Accidents	*Injures*	*Accidents*	*Deaths*	*Injures*
199s5	5.001	3.663	4.570	32	32	13
1996	5.018	3.766	4.604	29	29	6
1997	5.878	4.341	5.291	28	29	9
1998	5.581	4.322	5.343	36	37	22
1999	5.523	4.379	5.354	31	32	14
2000	5.619	4.273	5.208	28	29	7
2001	5.835	4.526	5.456	26	26	8
2002	5.577	4.388	5.325	23	23	10
Total	44.032	33.658	41.151	233	237	89
Mean	5.504	4.207	5.144	29	30	11

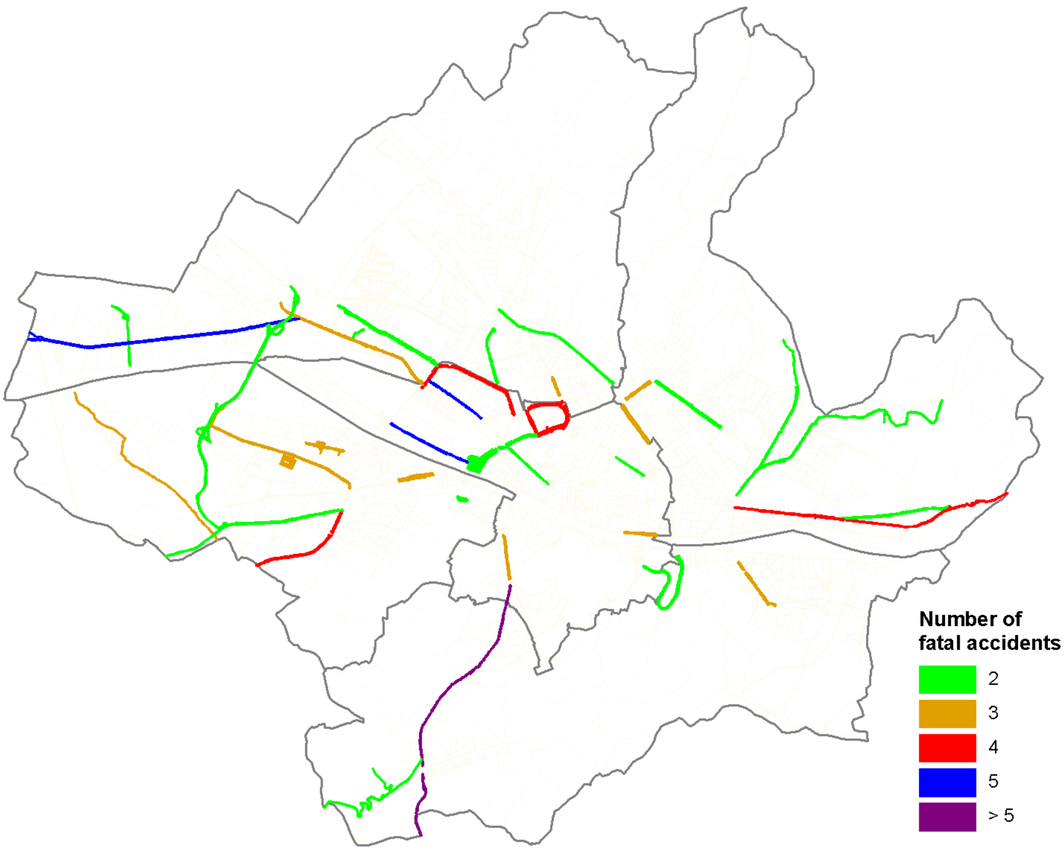
Table 2 *Index number: total accidents, fatal and non-fatal (with injures) accidents*

Index number	*Year*							
	1995	*1996*	*1997*	*1998*	*1999*	*2000*	*2001*	*2002*
Accidents	100	100	118	112	110	112	117	112
Accidents with injures	100	103	118	118	119	116	123	120
Fatal accidents	100	91	88	113	97	88	81	72

Considering table 2, we observe how both the total of accidents and those involving injuries grew quickly between 1996 and 1997, rising by about 20% with respect to 1995: at the contrary, fatal accidents show in these years a decreasing evolution, except an inversion trend in 1998 when the rise respect to 1995 is by 13%. As from 1999, this decrease has been going on constantly by about 10% each year. Between 1997 and 1999 the number of accidents was reduced, although slightly, whereas the number of accidents with injuries went on growing. This might have been due both to the major gravity of the accidents occurred in these years, and to the diminution of accidents with no injuries registered by the policemen that can be plausibly explained by the widespread use of bonus-malus insurance policies which prevent people from making a claim to avoid an increase in the insurance premium.

As for road accidents, the spatial component of the phenomenon is doubtlessly of major importance: the possibility to know where these events happen is expedient to undertake all the more fast, targeted and strategic intervention actions. Up until a few years ago the Geographical Information System required a deep knowledge in programming and considerable hardware cost for managing and processing a large amount of data. It has been spreading rapidly also in public offices, where the territory management and planning become more and more complex, and its applications offer timely and exhaustive responses to various requirements, from town planning to statistical and demographic purposes. The linkage between the database and the road digital map had to be modified to be effective, since in fact it was necessary to create new linkage keys to enable the two systems to interact.

Figure 5 *Roads with high number of fatal accidents – Florence (years 1995-2002)*

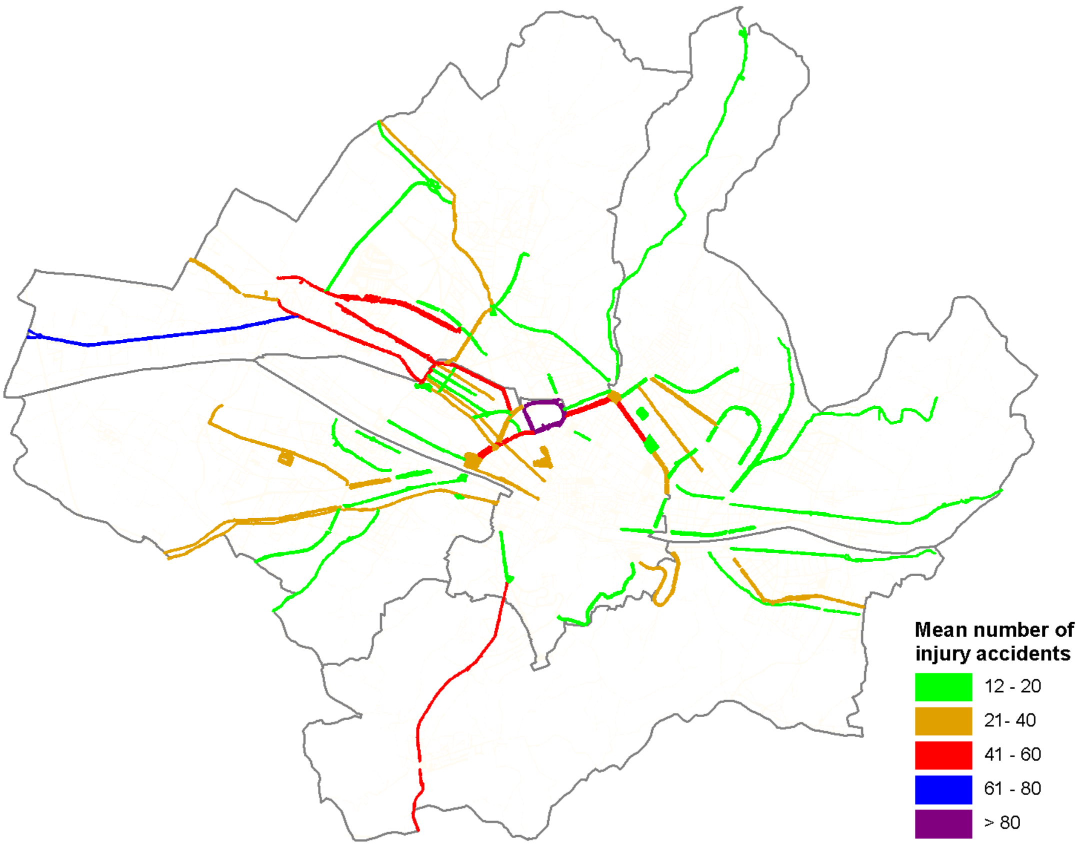


Once the connection established, it has been possible to achieve the complete mapping of the accidents in the municipal territory. The more interesting aspects, which better express the system's potentialities, are the possibility to test directly on the map every single accident to get full information on it, and the real time update of this information, allowing a permanent monitoring of the phenomenon, in order to characterize in time the critical points of the Florentine traffic.

With exclusive reference to accidents with personal injuries, distinguishing between accidents with death and accidents with physical damages, we achieved two lists of the roads where took place a high number of accidents in the reference period.

As for accidents involving only physical damages, in order to highlight the roads mainly concerned with the phenomenon, we limited our analysis to the areas of circulation that had in the reference period an average of at least one accident per week. We found that in the whole city there are 82 streets meeting this condition: Viale Filippo Strozzi, characterized by an annual average close to 100 accidents with injured, is the area of circulation where took place the higher number of accidents each year, except in 2002, standing out the other areas (Via Pistoiese, Via Senese, Viale Fratelli Rosselli, Via Baracca and Via di Novoli, the annual average of which ranks between 50 and 70 accidents per year, equal to more than one per week), as shows the maps in Figure 5 and 6.

Figure 6 *Roads with high number of injury accidents – Florence (years 1995-2002)*



Overall, accidents happened in 159 streets and excluding Via Senese, the annual average of lethal accidents is always less than 1. The map in Figure 3 helps find the streets mainly concerned with these tragic events, showing the areas of circulation where there were at least two of them in the reference period: there are 37 of them, representing 47% of lethal accidents all over the city.

4. Discussion and conclusions

The results of this work are meant as a contribution to the achievement of a coherent and shared information system to deal with the mobility phenomenon.
Therefore, we first focus our attention on the census data and its use, since censuses represent the occasion to harmonize languages, nomenclatures, classifications and codings relating not only to the surveyed phenomena but also to the territory, the key issue for measuring and analysing mobility.

In fact the “territory” constitutes, for the analysis of this phenomenon, the strategic sphere to which refers the logic of combining sources of different natures (statistical and administrative, systematic and/or occasional), logic that must lead to the achievement of a relational system between these sources. This is the sense of the two case-studies, that make reference to the municipal territory not, obviously, as the only area to be considered when dealing with mobility, but as the more characteristic of the management aspects of this phenomenon.
In the further phases of the study, the accurate analysis of the data quality will allow to make elaborations necessary to evaluate the correlations between mobility and traffic volumes, flows, speed, circulating fleet composition, accident density and their causes.
It will thus be possible to implement more complex statistical models capable of analysing both mobility and road accident phenomena, the latter in terms of dangerousness of the municipal road network.
The results contained in the present work constitute the starting point to study in depth the mobility phenomenon. The chosen study application field, the municipal area, is at the time being due to the available data, but referring to a larger territory area will allow both to study appropriate statistical models for the mobility analysis and to experiment models that explicitly take into account the special characteristics of road security.

References

Alidori G. (2003) Gli incidenti stradali a Firenze 1995-2002. La statistica per la città. Comune di Firenze.

Bailey T. C., Gatrell A. C. (1995) *Interactive Spatial Data Analysis*. Longman, London.

Cliff A. D. and Ord J. K,. (1973) *Spatial Processes: Models and Applications*. Pion, London

Fotheringham A.S. (1983) Some theoretical aspects of destination choice and their relevance to production-constrained gravity models. *Environment and Planning A* 15: 1121 – 1132

Fotheringham A.S. (1986) Modelling hierarchical destination choice. *Environment and Planning A* 18: 401–418

Getis A. (1991) Spatial interaction and spatial autocorrelation: A cross-product approach. *Environment and Planning A*: 1269–1277

Haining R. P. (1978) Estimating spatial interaction models. *Environment and Planning A* 10: 305-320

Huggett R. (1980) *Modelling in geography*, Harper & Row, London.

Petrucci A., Salvati N. (2003). "Un modello di interazione spaziale per l'analisi dei flussi a livello comunale". Atti del Convegno Intermedio della Società Italiana di Statistica su "*Analisi Statistica Multivariata per le Scienze Econimico-Sociali, le Scienze Ambientali e la Tecnologia*", Napoli 9-11 giugno 2003 (*in italian*)

Roy JR (2004) *Spatial interaction modelling: A regional science context.* Springer, Berlin Heidelberg NewYork

Stillwell J. (1991) Spatial interaction models and the propensity to migrate over distance, in: *Migration Models: Macro and Micro Approaches*, Stillwell J., Congdon P. (Eds.), Belhaven Press, London.

Wilson AG(1967) A statistical theory of spatial distribution models. *Transportation Research* 1: 253–269